

EXHIBIT 1

SUSMAN GODFREY L.L.P.

A REGISTERED LIMITED LIABILITY PARTNERSHIP
ONE MANHATTAN WEST
NEW YORK, NEW YORK 10001-8602
(212) 336-8330
FAX (212) 336-8340
WWW.SUSMANGODFREY.COM

SUITE 5100
1000 LOUISIANA STREET
HOUSTON, TEXAS 77002-5096
(713) 651-9366

SUITE 1400
1900 AVENUE OF THE STARS
LOS ANGELES, CALIFORNIA 90067-6029
(310) 789-3100

SUITE 3000
401 UNION STREET
SEATTLE, WASHINGTON 98101-2683
(206) 516-3880

ZACH SAVAGE
DIRECT DIAL (212) 729-2022

E-MAIL ZSAVAGE@SUSMANGODFREY.COM

November 8, 2024

VIA ELECTRONIC MAIL

Vera Ranieri
Morrison & Foerster LLP
vranieri@mofo.com
425 Market Street
San Francisco, CA 94105-2482

Elana Nightingale Dawson
Latham & Watkins LLP
elana.nightingaledawson@lw.com
505 Montgomery Street, Suite 2000
San Francisco, CA 94111

Katie Lynn Joyce
Keker, Van Nest & Peters LLP
kjoyce@keker.com
633 Battery Street
San Francisco, California 94111-1809

Re: *The New York Times Company v. Microsoft Corporation, et al.*,
Case No.: 23-cv-11195-SHS

Counsel:

I write regarding OpenAI's Responses and Objections to The Times's Second Set of Requests for Production of Documents, served on July 12, 2024, and OpenAI's Responses and Objections to The Times's Third Set of Requests for Production of Documents, served on September 9, 2024. This letter follows up on the parties' meet and confers about these Sets of Requests, including videoconferences on August 26, August 29, October 18, and October 24, as

well as the parties' written correspondence about these Requests, including July 31 and August 20 letters, and email correspondence on October 4, October 11, October 15, October 28, and November 1.

Consistent with the Court's instructions during the October 30 hearing, and the Court's November 1 Order, The Times intends to file any motion concerning unresolved disputes by November 19, 2024. The Times is available to meet and confer on November 12, November 13, and/or November 14, and hopes to resolve these disputes without the need for motion practice.

I. OpenAI Responses that Should Be Updated Based on Related Agreements

For many RFPs, OpenAI previously limited its responses to specific models and products, namely, the "models used for ChatGPT", and/or "ChatGPT" itself.

The parties have now agreed that OpenAI will produce non-custodial documents concerning GPT-2 and GPT-3, "as [OpenAI] has for other models," and that OpenAI will produce custodial documents for GPT, GPT-2, and GPT-3. Dkt. 262 at 2. OpenAI has also agreed to produce documents concerning additional products. *See* Dkt. 302 (OpenAI confirming that the following products are now subject to discovery "ChatGPT (including Browse (f/k/a Browse with Bing)), the OpenAI API, SearchGPT, and Microsoft's Copilot," as well as "the specific third-party Custom GPTs referenced in Plaintiffs' Complaints and in their Motion").

Please confirm that OpenAI will revise its RFP responses accordingly, including for at least RFPs 24, 30, 34, 39, 43, 49, 61, 72, and 89. These are responses where OpenAI previously limited its productions to documents concerning the "models used for ChatGPT" and/or "ChatGPT" itself. These responses and productions should now be updated in light of parties' agreements about the additional models and products subject to discovery.

II. Outstanding Document Disputes

The Times below summarizes the current status of several disputed RFPs.

RFP 2 ("Documents You have gathered to submit or have submitted to any legislative or executive agency, committee, or other governmental entity in the United States that concern or relate to the allegations in the Complaint")

After the September 12 hearing, OpenAI produced limited Interrogatory responses that OpenAI had provided in response to the FTC's Civil Investigative Demand No. 232-3044. *See* OPCO_NEWS_0621960; OPCO_NEWS_0622046. On October 11, we emailed you seeking additional responses to the same FTC Demand. Specifically, we requested that OpenAI produce its responses to FTC Interrogatory Nos. 2, 3, 5, 6, 7, and 10. OpenAI has not yet responded to our request. Please confirm whether OpenAI agrees to produce these additional responses.

In addition, OpenAI should produce documents related to its discussions with the California Attorney General concerning OpenAI's corporate structure and nonprofit status. *See*

<https://www.reuters.com/technology/openai-talks-with-california-become-for-profit-company-bloomberg-news-reports-2024-11-04/>. Please confirm whether OpenAI will do so.

We also ask that OpenAI be prepared to discuss during our meet and confer other proceedings in which OpenAI has submitted documents relevant to the issues in this case, so we can evaluate the proper scope for OpenAI's productions in response to this RFP.

RFP 16 ("Documents concerning the decision not to release Training Datasets or other information for Defendants' Text Generation AI Model(s) after GPT-3 (including all versions thereof, such as GPT-4 Turbo and GPT-4o")

OpenAI has so far refused to produce any documents in response to this Request, which is relevant to, among other issues, OpenAI's anticipated fair use defense. Please confirm whether OpenAI will produce the requested documents.

RFP 17 ("Documents concerning the compilation, acquisition, and curation of Training Datasets including: (i) communications regarding paying for or obtaining permission to use any Training Datasets or content therein, ii) communications with Training Dataset creators, including Common Crawl, and iii) communications concerning the relevance and impact of paywalls on building Training Datasets or gathering training data.")

The parties' related discussions about RFP 6 have narrowed the scope of the dispute. But there remains a dispute because OpenAI has not clarified whether it is producing communications with Training Dataset creators, including Common Crawl. Please confirm OpenAI is doing so.

Please also confirm that the parties' agreements regarding other RFPs will cover item (iii) from this Request (regarding paywalls), including RFPs 1, 3, 44 and 53.

RFP 23 ("Documents concerning all steps taken or contemplated to train, instruct, program, or refine Defendants' Generative AI Products and Services using Times Content or Journalism content")

The Times accepts OpenAI's offer to resolve this RFP, as memorialized in Max Levy's October 15 email, subject of course to the parties' global dispute about the relevant products.

"OpenAI also agrees to conduct a reasonable search for and produce non-privileged documents in its possession, custody, or control concerning (i) how and why OpenAI curates and uses different types of data to train OpenAI's relevant models; and (ii) the relative value of types of data for use in the relevant models. Further, OpenAI will conduct a reasonable search for and produce any non-privileged documents in its possession, custody, or control describing whether and how the relevant models were trained using articles published by The New York Times. Finally, OpenAI will conduct a reasonable search for and produce any non-privileged

documents in its possession, custody, or control sufficient to show OpenAI's understanding of how the models used for ChatGPT learn from the data used to train the model."

RFP 24 ("Documents concerning any analyses, studies, measurements, testing, experimentation, assessments, or other evaluation of the incremental value of including Journalism content, including Times Content, in Defendants' Training Datasets")

OpenAI has so far only agreed to produce documents "sufficient to show OpenAI's evaluation or assessment of the value of including Journalism content in the training datasets used to train the models used for ChatGPT." Given that the use of Journalism to train Defendants' generative AI models is a central issue in the case, a "sufficient to show" limitation is improper for this request. OpenAI should also produce custodial documents found pursuant to a reasonable search. Further, OpenAI should withdraw its limitation to documents concerning "models used for ChatGPT." Please confirm OpenAI will do so.

RFP 28 ("Documents concerning all measures taken to conceal or obscure Defendants' use of Times Content during any stage of the development of each of Defendants' Generative AI Products and Services")

OpenAI on October 15 offered to produce "any non-privileged documents in its possession, custody, or control concerning whether to remove from datasets on which the relevant models were trained the following categories of information with respect to copyrighted works: authorship, date of publication, publishing entity, title, copyright notices, and terms and conditions for use."

The Times's counter-proposal is as follows. For both model training and RAG, please produce documents concerning the removal or alteration, if any, of copyright management information, including: information regarding authorship, date of publication, publishing entity, title, copyright notices, and terms and conditions for use. This aspect of the counter-proposal would also resolve the parties' pending dispute over RFP 18. See Dkt. 283. In addition, OpenAI should produce documents concerning whether OpenAI intended, planned, or made any effort to conceal or withhold from disclosure the source of its training data or the source of content used for RAG. The addition of "whether" to this sentence should resolve any concerns about the word "conceal," while also clarifying that we seek documents regarding any withholding of this information even if OpenAI would not characterize it as concealment.

RFP 35 (“Documents concerning any changes made to Defendants’ Generative AI Models or Generative AI Products and Services from 2021 to present, including any logs showing changes, documents relating to ablation or ‘unlearning,’ (i.e. the ability to remove memorization, content, or other effects of specific parts of the Training Datasets from the models), and documents and communications concerning any changes to the models made or considered, including in response to claims of copyright infringement or improper use of data.”).

The Times is willing to narrow this Request to the portion seeking documents and communications concerning any changes to the models made or considered in relation to ablation or unlearning or in response to claims of copyright infringement or improper use of data. Please confirm OpenAI will produce documents in response to this narrowed version of the Request.

RFPs 39-41: These Requests seek documents concerning problematic behaviors displayed by Defendants’ products and services. *See RFP 39 (“Documents concerning the study, testing, or evaluation of Defendants’ Generative AI Products and Services and Generative AI Models for hallucinations, memorization, and any problematic behaviors”); RFP 40 (“Documents concerning the prevalence of outputs that are verbatim copies of text included in Training Datasets”); RFP 41 (“Documents or surveys related to users’ trust in the accuracy or veracity of information contained in outputs of Defendants’ Generative AI Products and Services and how users respond to disclaimers regarding the accuracy or veracity of information contained in such outputs”).*

These Requests are relevant to, among other issues, The Times’s misappropriation claim, trademark dilution claim, and The Times’s harm allegations. OpenAI has so far refused to produce any documents in response to these Requests. Please confirm whether OpenAI will do so.

RFP 42 (“Documents concerning Defendants’ efforts and abilities to monitor, log, analyze, quantify, assess, evaluate, or review user queries within any Generative AI Products and Service”).

OpenAI has so far only agreed to produce “documents in its possession, custody, and control sufficient to show OpenAI’s efforts that are responsive to this request.” Please confirm this production will include information about OpenAI’s retention policies for this information and data.

In addition, why did OpenAI exclude the word “abilities” from its response? OpenAI should produce documents concerning its capabilities, in addition to any efforts specifically undertaken. OpenAI should also produce custodial documents in response to this Request. The parties are separately negotiating OpenAI’s production of log data, but those negotiations should not prevent OpenAI from collecting and producing responsive documents. To the contrary, such document productions will inform the parties’ discussions about data productions. We understand OpenAI will be providing an update by November 8 on its investigations into data productions. Based on those investigations, please confirm whether OpenAI will produce documents in response to this Request.

RFP 43 (“Any surveys, studies, analyses, or statistics on patterns and practices any of Defendants’ Generative AI Products and Services, including any mentions of The New York Times, such surveys, studies, analyses, or statistics regarding user engagement within user queries and/or inclusion of links to Times Content in user queries or outputs”)

OpenAI agreed to produce “documents in its possession, custody, and control sufficient to show OpenAI’s works that are responsive to this request.” OpenAI later clarified in its August 20 letter that the term “OpenAI work” refers to any “surveys, studies, analyses, or statistics on patterns and practices” conducted, generated, or commissioned by OpenAI relating to ChatGPT.” OpenAI also clarified in the same letter that “to the extent such works were prepared by an entity other than OpenAI but is within OpenAI’s possession, custody or control, they would be included.”

As noted above, *supra* Section I, these productions should now include documents concerning the additional products that OpenAI has agreed will be subject to discovery. The only other remaining disagreement concerns whether OpenAI will produce custodial documents in response to this Request. Please confirm OpenAI will do so.

RFPs 45, 52: These Requests seek information about OpenAI logs. *See RFP 45* (“Query, session, and chat logs reflecting or analyzing user sessions related to Times Content within Defendants’ Generative AI Products and Services, including, for each session, user queries paired with responses to those queries”); *RFP 52* (“Documents concerning the extent to which Defendants’ Generative AI Products and Services direct users to Times Content in response to user queries, including by providing outputs that directly quote Times Content, contain a URL that is intended to direct the user to Times Content, or reference The Times or one of its contributors, authors, contractors, or employees”).

OpenAI on November 1 took the position that “responsive information for both these RFPs will be contained in logs, which are the subject of [an] ongoing discussion” between Jen Maisel and Paven Malhotra. But that does not answer whether OpenAI also has documents responsive to these Requests. Does it?

RFP 47 (“Documents concerning user engagement metrics, events, and analytics data collected during each user session, including duration, the number of prompts or queries that a user inputs per session, click-through data, fallback rates, and bounce rates”).

The parties are close to an agreement on this Request. OpenAI has as of October 24 agreed “to produce responsive, non-privileged documents in its possession, custody, and control, if any, concerning the following user engagement metrics: time-spent, average number of prompts, click-through data, and bounce rate.”

The remaining dispute is whether OpenAI will also produce documents concerning fallback rates. Our October 28 email provided more information about fallback rates, which are relevant to, among other issues, The Times’s misappropriation claim. This discovery addresses

The Times's allegation that Defendants' products usurp commercial opportunities for The Times, including for Wirecutter.

Please confirm OpenAI will produce the requested documents.

RFP 49 ("Documents concerning Defendants' monitoring, tracking, and knowledge of investigations into whether Defendants' AI Models contain copyrighted content, including Times Content.").

OpenAI initially only agreed to produce documents "sufficient to show OpenAI's knowledge, prior to the inception of this lawsuit, of whether New York Times content was included in a dataset used for any model used for ChatGPT." OpenAI on August 20 later agreed to remove the "sufficient to show" limitation. Please confirm that by withdrawing this limitation, OpenAI is agreeing to produce custodial documents in response to this request.

Assuming that is the case, there are two remaining issues. First, OpenAI should also produce documents concerning its knowledge of any copyrighted content (not just Times content), including Daily News Plaintiffs' content and the Center for Investigative Reporting's content. Second, the time period should not be limited to "prior to the inception of this lawsuit." Please confirm that OpenAI will withdraw these two limitations.

RFP 54 ("Documents concerning any public statements made by You or Your employees relating to this litigation, including: (i) Your claim that the New York Times 'hacked' Defendants' Generative AI Products, ii) Your contention that Times Content is 'not significant for the model's intended learning' and that it did not 'meaningfully contribute to the training of [Your] existing models,' iii) any contention by You that You can modify Defendants' Text Generation AI Models to minimize the 'regurgitation' of copyrighted content, and iv) Your efforts to use artificial intelligence to 'monetize news and other published content.'")

Your October 15 email offered to produce documents regarding certain public statements that we identified in our October 4 email. But OpenAI still refuses to produce documents concerning the specific statements mentioned in the RFP that were included in OpenAI's January 8, 2024 publication titled "OpenAI and Journalism." Specifically, OpenAI is refusing to produce documents concerning statements (ii) and (iii) above. We understand that the parties are at an impasse on these aspects of the Request.

RFP 57 ("Documents related to the use of Defendants' Generative AI Products and Services, including by third parties, to promulgate misinformation, disinformation, or poor quality 'pinkslime' journalism")

OpenAI has so far refused to produce any documents in response to this Request, which is relevant to, among other issues, The Times's harm allegations. To aid your investigation, we can point you to this News Guard study which discusses "AI-generated news and information sites

operating with little to no human oversight.” <https://www.newsguardtech.com/special-reports/ai-tracking-center/>. Please confirm OpenAI will produce documents in response to this Request.

RFP 58 (“Documents concerning Your review and approval process for 1) custom GPTs made available in Your Custom GPT Store, and 2) third-party ChatGPT plug-ins made available in Your ChatGPT plug-in store, including custom GPTs and plug-ins designed to remove paywalls, provide Journalism, or that otherwise reference The Times or Journalism.”)

OpenAI responded that it will only produce documents “sufficient to show OpenAI’s review and approval process for the GPT Store and third-party ChatGPT plug-ins that it locates pursuant to a reasonable and diligent search.” Please confirm that OpenAI’s production will include documents sufficient to show its review and approval process for custom GPTs and plug-ins designed to remove paywalls and/or provide Journalism.

RFP 60 (“Documents concerning Defendants’ licensing or use of Bing’s search index”)

OpenAI has taken the position that it will not produce any documents in response to this Request because “it has not used Bing Index data to train any GPT model used for ChatGPT.” As you know, this case is not limited to training, but instead involves the use of Times content in Defendants’ generative search products as well. *See* Compl. ¶ 108. Does OpenAI use the Bing Index to operate any generative AI products? Also, has OpenAI used Bing Index data to train other models, which are now in scope given the parties’ separate agreement with respect to the motion to compel filed at Dkt. 128 (dispute regarding GPT, GPT-2, and GPT-3)? We asked these questions in our July 31, 2024 letter, and OpenAI has still not answered them. Please do so.

RFP 64 (“Documents concerning competition and comparisons between Defendants’ Generative AI Products and Services and those of other companies, including (i) how Generative AI tools enable You to compete in search, (ii) how Defendants’ use of copyrighted materials can help You compete with Google, (iii) any analyses and projections of how increasing Your search market share will boost Your revenue, (iv) the benefits of being a search or knowledge “destination,” and (v) Google’s reduction of search referral traffic (via SGE and direct answers) and stated goal of keeping users on its site.”).

OpenAI has so far refused to produce any documents in response to this Request, which is relevant to, among other issues, fair use factor four, quantifying OpenAI’s unjust enrichment, and harm to The Times. The Times stands by this Request and requests that OpenAI produce responsive documents. Please confirm whether OpenAI will do so.

RFP 66 (“Documents concerning fundraising efforts related to Defendants’ Generative AI Products and Services, including any investor solicitation efforts and any solicitation of charitable or in-kind contributions.”)

Our October 4 email offered to narrow this Request to “documents relating to (i) financial data, including projections; (ii) OpenAI’s plans and strategies to commercialize and/or monetize its products; (iii) OpenAI’s non-profit structure; (iv) OpenAI’s use of copyrighted content; and (v) potential or existing lawsuits or investigations that are related to any of the issues in the Complaint.”

Your October 15 email stated that “OpenAI has already agreed to conduct a reasonable search for non-privileged documents in its possession, custody, or control addressing most of the substantive issues your letter identifies. *See, e.g.*, RFP Nos. 70-71 (seeking documents related to certain financial data), RFP Nos. 10, 15, 68 (seeking documents related to commercialization of OpenAI’s products), RFP No. 14 (seeking documents concerning OpenAI’s non-profit structure), RFP No. 3 (seeking documents concerning the use of copyrighted material), and RFP No. 4 (seeking documents concerning claims of copyright and trademark infringement).”

We interpret your October 15 email to mean that OpenAI has already agreed to produce the documents requested in items (i), (ii), (iii), and (iv) from our October 4 email. Please confirm.

OpenAI, however, still refuses to produce documents regarding item (v). The parties are at an impasse as to this aspect of the Request.

RFPs 67, 70, 71, 72, 88: These Requests seek financial information that is relevant to, at a minimum, The Times’s disgorgement damages. *See* RFP 67 (“Documents sufficient to show (i) projected revenue streams from Defendants’ Generative AI Products, (ii) past, present, and anticipated future subscriber numbers, and (iii) product-level revenue data from 2021 to present”); RFP 70 (“Documents concerning referral, affiliate, advertising, and paid search revenue generated by Defendants’ Generative AI Products and Services.”); RFP 70 (“Documents sufficient to show in detail, on a monthly, quarterly, or annual basis, the revenue and profit realized from Defendants’ Generative AI Products and Services, including documents sufficient to show in detail, by product or service, the gross revenue generated, the costs, the gross profit margin, and the net profit margin.”); RFP 72 (“Documents relating to the pricing of Defendants’ Generative AI Products and Services in the United States, including price quotes, price lists, subscription amounts, price announcements, and actual prices paid by customers.”); RFP 88 (“Documents sufficient to show, on a monthly basis from December 2015 to present, 1) OpenAI’s valuation, 2) the number of employees of each OpenAI entity named in the Times’s Complaint, and 3) OpenAI’s capital investment, including the dollar amount of capital purchases, number of GPUs used, and other capital expenses to train AI models”).

OpenAI has so far refused to produce any documents in response to RFPs 67 and 88, which seek discrete information that OpenAI can easily locate and produce. Please confirm OpenAI will do so.

For RFPs 70 and 71, OpenAI has only agreed to produce documents “sufficient to show income from paid versions of ChatGPT.” Our July 31 letter asked a few questions about this limitation, but OpenAI has not yet responded to these questions. Instead, OpenAI in its August 20 letter directed us to one document, OPCO_NYT_0192848. But this document simply lists the “revenue” and “costs” for ChatGPT, without providing any of the more detailed information requested by these RFPs. Please respond to the questions in our July 31 letter, and in doing so, please keep in mind OpenAI’s recent agreement to expand discovery beyond just the ChatGPT product. *See* Dkt. 302.

For RFP 72, OpenAI simply provided a link to a website where it contends that “pricing for ChatGPT is available.” Is that the only responsive document? Has OpenAI reconsidered this response given its agreement that additional products are now in scope aside from just ChatGPT?

RFP 69 (“Documents concerning Your policies or practices regarding public access to Your governing documents, financial statements, and conflict of interest rules.”).

Our October 4 email offered to narrow this Request to documents “sufficient to show” why OpenAI in 2022 changed its practices regarding public access to OpenAI’s governing documents, financial statements, and conflict of interest rules, as reflected in this article: <https://www.wired.com/story/openai-scrapped-promise-disclose-key-documents/>. This Request is relevant to, among other issues, OpenAI’s anticipated fair use defense.

OpenAI on October 15 confirmed OpenAI is still refusing to provide any documents in response to this Request, even as narrowed. The parties are at an impasse.

RFP 82 (“All communications with third parties concerning Defendants’ fair use defense to copyright infringement, including how any licensing deals would impact that defense”).

OpenAI has so far refused to produce any documents in response to this Request, which is highly relevant to OpenAI’s anticipated fair use defense. Please confirm OpenAI will produce these documents.

RFPs 86-87: These RFPs seek documents concerning synthetic datasets. *See* RFP 86 (“All Documents concerning use of Generative AI Models to create Synthetic Datasets”); RFP 87 (“All Documents concerning Your use or contemplated use of Synthetic Datasets, including to train Your Generative AI Models.”).

OpenAI has so far refused to produce any documents in response to these Requests, which are relevant to The Times’s infringement claims. For example, any synthetic datasets created by OpenAI by using Times content may themselves be infringing works. OpenAI’s development of synthetic datasets is also relevant to fair use, including insofar as these synthetic datasets impact the market for training data.

Please confirm OpenAI will produce documents responsive to these Requests.

RFP 91 (“All minutes of meetings of the board of directors of OpenAI, Inc., OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC, and all Documents provided to those boards of directors that concern or relate to 1) Your training data 2) Your intention to commercialize or monetize Defendants’ Generative AI Models, Products, and Services, and 3) projections regarding revenues or profits for Your Generative AI Models, Products, or Services.”)

OpenAI has refused to produce any documents in response to this Request. That refusal is untenable, including because each aspect of this Request addresses topics for which OpenAI has otherwise agreed to produce documents. OpenAI cannot withhold a document from discovery merely because it was provided to a board of directors. Board minutes that address these topics are also relevant.

Please confirm OpenAI will produce documents in response to this Request.

RFP 95 (“Documents and communications concerning Defendants’ [REDACTED] referenced in OPCO_NYT_0000668 and OPCO_NYT_0000669.”)

This Request targets discrete but highly relevant information identified through OpenAI’s prior productions. The cited documents describe [REDACTED]

[REDACTED] Please confirm OpenAI will produce the requested documents.

Respectfully,

/s/ Zach Savage
Zach Savage